

# ANALYSE DISCRIMINANTE MATRICIELLE DESCRIPTIVE. APPLICATION À L'ÉTUDE DE SIGNAUX EEG

Juliette Spinnato<sup>1,3,4</sup> & Marie-Christine Roubaud<sup>1</sup> & Margaux Perrin<sup>2</sup> & Emmanuel Maby<sup>2</sup> & Jeremie Mattout<sup>2</sup> & Boris Burle<sup>3</sup> & Bruno Torr sani<sup>1</sup>

<sup>1</sup> Aix-Marseille Universit , CNRS, Centrale Marseille, I2M, UMR 7373, 13453 Marseille

<sup>2</sup> Universit  Lyon 1, CNRS, INSERM U1028, CRNL, UMR 5292, 69000 Lyon

<sup>3</sup> Aix-Marseille Universit , CNRS, LNC, UMR 7291, 13331 Marseille

<sup>4</sup> juliette.spinnato@univ-amu.fr

**R sum .** Nous nous int ressons   l'approche descriptive de l'analyse discriminante lin aire de donn es matricielles dans le cas binaire. Sous l'hypoth se de s parabilit  de la variabilit  des lignes de celle des colonnes, les combinaisons lin aires des lignes et des colonnes *les plus discriminantes* sont d termin es par la d composition en valeurs singuli res de la diff rence des moyennes des deux classes en munissant les espaces des lignes et des colonnes de la m trique de Mahalanobis. Cette approche permet d'obtenir des repr sentations des donn es dans des plans factoriels et de d gager des composantes discriminantes. Une application   des signaux d' lectroenc phalographie multi-capteurs illustre la pertinence de la m thode.

**Mots-cl s.** Analyse discriminante lin aire, donn es matricielles, matrice de covariance s parable, d composition en valeurs singuli res, signaux EEG

**Abstract.** We focus on the descriptive approach to linear discriminant analysis for matrix-variate data in the binary case. Under a separability assumption on row and column variability, the *most discriminant* linear combinations of rows and columns are determined by the singular value decomposition of the difference of the class-averages with the Mahalanobis metric in the row and column spaces. This approach provides data representations of data in two-dimensional or three-dimensional plots and singles out discriminant components. An application to electroencephalographic multi-sensor signals illustrates the relevance of the method.

**Keywords.** Linear discriminant analysis, matrix-variate data, separable covariance matrix, singular value decomposition, EEG signal

## 1 Introduction

L'analyse discriminante lin aire (LDA) de donn es structur es de type matriciel est  tudi e dans de nombreux contextes notamment en classification de signaux multi-capteurs telles que les donn es  lectroenc phalographiques (EEG) (e.g. Guit rrez & Escalona-Vargas, 2010). Une approche simple pourrait  tre de vectoriser les donn es et d'utiliser la LDA classique sans tenir compte de leur structure. Cependant, les vecteurs ainsi construits

sont généralement de très grande dimension et la LDA classique est mise à défaut. En effet, le nombre de paramètres à estimer est souvent supérieur au nombre d'observations et l'estimation de la matrice de covariance est donc de très mauvaise qualité. Afin de réduire le nombre de paramètres à estimer il est usuel d'introduire une hypothèse de séparabilité entre la variabilité des lignes et celle des colonnes. Ceci permet de prendre en compte la structure des données par un modèle de covariance simple sous la forme du produit de Kronecker de la matrice de covariance des lignes et de celle des colonnes (e.g. Akdemir & Gupta, (2011) et les références citées). Cette hypothèse est raisonnable dans plusieurs situations et notamment dans l'étude des signaux EEG, cadre applicatif de ce travail (e.g. Bijma *et al.*, (2005), Friston *et al.*, (2005) et Mahanta *et al.*, (2012)).

L'objectif de ce travail est de proposer une méthode de détermination de combinaisons linéaires de lignes et de colonnes *les plus discriminantes* entre deux classes en tenant compte de la structure des données sous l'hypothèse de séparabilité. Notre approche repose sur la décomposition en valeurs singulières (SVD) de la différence des moyennes en munissant les espaces des lignes et des colonnes de la métrique de Mahalanobis. Ceci nous permet d'extraire conjointement les combinaisons *les plus discriminantes* dans ces deux espaces et d'en déduire des représentations graphiques de ces composantes ainsi que des visualisations des données matricielles dans des sous-espaces de dimension réduite. Une approche similaire a déjà été appliquée à l'extraction de caractéristiques discriminantes des potentiels d'erreur dans les signaux EEG dans un contexte de classification binaire par Spinnato *et al.* (2014). Dans ce travail nous développons l'aspect descriptif de la méthode et étudions son apport pour des signaux EEG dans un paradigme d'Interface Cerveau-Machine de type P300 Speller (Farwell & Donchin, 1988).

La structure de cet article est la suivante. Nous introduirons tout d'abord quelques notations et définitions. Nous présenterons ensuite l'approche matricielle descriptive de la LDA binaire sous l'hypothèse de séparabilité et son apport en termes de réduction de dimension et d'extraction de caractéristiques discriminantes. Enfin nous montrerons la pertinence de la méthode pour l'étude des caractéristiques discriminantes de potentiels évoqués en EEG.

**Notations et définitions.** Notons  $\mathbb{R}^{K \times J}$  l'espace des matrices réelles à  $K$  lignes et  $J$  colonnes. Soient  $M \in \mathbb{R}^{J \times J}$  et  $D \in \mathbb{R}^{K \times K}$  deux matrices symétriques et définies positives. L'espace des lignes  $\mathbb{R}^J$  est dit muni de la métrique  $M$  si le produit scalaire sur cet espace est défini par  $\langle x, y \rangle_M = x'My$  pour tout  $x, y \in \mathbb{R}^J$ . De manière similaire, l'espace des colonnes  $\mathbb{R}^K$  est muni de la métrique  $D$ .

**Définition 1** (Produit scalaire, norme matricielle). Soient  $X$  et  $Y$  des matrices  $\in \mathbb{R}^{K \times J}$ ,  

$$\langle X, Y \rangle_{M,D} = \text{Tr}(XMY'D) \text{ et } \|X\|_{M,D}^2 = \text{Tr}(XMX'D), \quad (1)$$

où  $\text{Tr}$  désigne l'opérateur trace et  $X'$  la matrice transposée de  $X$ .

Sachant que  $\text{Tr}(A'BCD) = \text{vec}(A)'(D \otimes B)\text{vec}(C)$ , où  $\text{vec}(A)$  désigne la vectorisation de la matrice  $A$  par concaténation de ses colonnes et  $\otimes$  le produit de Kronecker, on obtient

l'équivalence entre le produit scalaire défini en (1) sur  $\mathbb{R}^{K \times J}$  et celui défini sur l'espace  $\mathbb{R}^{KJ}$  muni de la métrique  $D \otimes M$ :

$$\langle X, Y \rangle_{M,D} = \text{vec}(X')'(D \otimes M)\text{vec}(Y') = \langle \text{vec}(X'), \text{vec}(Y') \rangle_{D \otimes M}. \quad (2)$$

## 2 Analyse discriminante matricielle descriptive sous l'hypothèse de séparabilité

### Modèle et estimation

Nous nous intéressons à l'analyse discriminante binaire dans le cas de données structurées sous forme matricielle. Dans chacune des deux classes  $c = 1, 2$ , les observations  $X_c^i \in \mathbb{R}^{K \times J}$  pour  $i = 1, \dots, n$ , sont considérées comme des réalisations i.i.d. d'une matrice aléatoire  $X_c$  de moyenne  $\mu_c$  dépendante de la classe. De plus, nous supposons que les matrices de covariance des lignes  $\Sigma_L \in \mathbb{R}^{K \times K}$  et des colonnes  $\Sigma_R \in \mathbb{R}^{J \times J}$  sont indépendantes de  $c$ , inversibles et telles que

$$\Sigma = \Sigma_L \otimes \Sigma_R, \quad (3)$$

où  $\Sigma \in \mathbb{R}^{KJ \times KJ}$  est la matrice de covariance de  $\text{vec}(X'_c) \in \mathbb{R}^{KJ}$  correspondant à la vectorisation de la matrice  $X'_c$ .

Soient  $n_1$  et  $n_2$  les effectifs des classes 1 et 2. Posons  $n = n_1 + n_2$ . Comme définie par Dutilleul P. (1999), l'estimation des matrices de covariance  $\Sigma_L$  et  $\Sigma_R$  s'effectue de manière itérative via une normalisation des  $J$  colonnes et des  $K$  lignes:

$$S_W^L = \frac{1}{nJ} \sum_{c=1}^2 \sum_{i=1}^{n_c} (X_c^i - \bar{X}_c)(S_W^R)^{-1}(X_c^i - \bar{X}_c)', \quad (4)$$

$$S_W^R = \frac{1}{nK} \sum_{c=1}^2 \sum_{i=1}^{n_c} (X_c^i - \bar{X}_c)'(S_W^L)^{-1}(X_c^i - \bar{X}_c), \quad (5)$$

où  $\bar{X}_c = \frac{1}{n_c} \sum_{i=1}^{n_c} X_c^i$ . Nous en déduisons l'estimation de la matrice de covariance  $\Sigma$  :

$$S_W = S_W^L \otimes S_W^R. \quad (6)$$

**Remarque 1** (Identifiabilité). *Pour tout  $\kappa \neq 0$ , notons que  $S_W^L \otimes S_W^R = \kappa S_W^L \otimes \kappa^{-1} S_W^R$ . Afin de résoudre ce problème d'identifiabilité la norme de  $S_W^R$  est fixée à 1 dans l'estimation.*

### Décomposition en valeurs singulières de la différence des moyennes

Rechercher les combinaisons linéaires des lignes et des colonnes les plus discriminantes équivaut à décomposer en valeurs singulières la différence des moyennes empiriques des deux classes en munissant  $\mathbb{R}^J$  l'espace des lignes (resp.  $\mathbb{R}^K$ , l'espace des colonnes) de la métrique  $M = (S_W^R)^{-1}$  (resp.  $D = (S_W^L)^{-1}$ ).

**Proposition 1.** Soit  $Q$  le rang de la matrice  $\bar{X}_1 - \bar{X}_2 \in \mathbb{R}^{K \times J}$ .

$$\bar{X}_1 - \bar{X}_2 = U \Lambda^{\frac{1}{2}} V' = \sum_{q=1}^Q \sqrt{\lambda_q} u_q v_q', \quad \text{où} \quad (7)$$

- $U \in \mathbb{R}^{K \times Q}$  a pour colonnes les vecteurs propres  $D$ -orthonormés de la matrice  $D$ -symétrique semi-définie positive  $(\bar{X}_1 - \bar{X}_2)M(\bar{X}_1 - \bar{X}_2)'D$  associés aux  $Q$  valeurs propres non nulles  $\lambda_q$  rangées dans un ordre décroissant dans la matrice diagonale  $\Lambda$ .
- $V \in \mathbb{R}^{J \times Q}$  a pour colonnes les vecteurs propres  $M$ -orthonormés de la matrice  $M$ -symétrique semi-définie positive  $(\bar{X}_1 - \bar{X}_2)'D(\bar{X}_1 - \bar{X}_2)M$  associés aux valeurs propres non nulles  $\lambda_q$  avec

$$V = (\bar{X}_1 - \bar{X}_2)' D U \Lambda^{-\frac{1}{2}}. \quad (8)$$

D'après (2), l'espace  $\mathbb{R}^{KJ}$  est muni de la métrique  $D \otimes M = S_W^{-1}$ . A partir de (7), sachant que  $\text{vec}(u_q v_q') = u_q \otimes v_q$ , nous obtenons une décomposition pour la différence des moyennes vectorisées et pour la distance de Mahalanobis entre ces moyennes :

**Proposition 2.** Soient  $u_q \otimes v_q$ ,  $q = 1, \dots, Q$ , des vecteurs  $S_W^{-1}$ -orthonormés de  $\mathbb{R}^{KJ}$ . Alors

$$\text{vec}(\bar{X}_1' - \bar{X}_2') = \sum_{q=1}^Q \sqrt{\lambda_q} (u_q \otimes v_q), \quad \text{et} \quad (9)$$

$$\|\text{vec}(\bar{X}_1' - \bar{X}_2')\|_{S_W^{-1}}^2 = \sum_{q=1}^Q \lambda_q \quad \text{avec} \quad \lambda_q = |\langle \text{vec}(\bar{X}_1' - \bar{X}_2'), u_q \otimes v_q \rangle_{S_W^{-1}}|^2. \quad (10)$$

**Projection dans l'espace lignes-colonnes  $\mathbb{R}^{KJ}$ .** Les coordonnées des observations  $X^i$  vectorisées sur les axes engendrés par les vecteurs  $u_q \otimes v_q$  pour  $q = 1, \dots, Q$  sont données par  $\langle \text{vec}((X^i)'), u_q \otimes v_q \rangle_{S_W^{-1}}$ . Pour la différence de moyennes, on obtient de plus l'erreur d'approximation.

**Proposition 3.** Soit  $\text{proj}_{E_r}(\text{vec}(\bar{X}_1' - \bar{X}_2')) = \sum_{q=1}^r \sqrt{\lambda_q} (u_q \otimes v_q)$  la projection de la différence des moyennes sur  $E_r$  le sous-espace engendré par les  $r$  premiers vecteurs  $u_q \otimes v_q$ .

$$\|\text{vec}(\bar{X}_1' - \bar{X}_2') - \text{proj}_{E_r}(\text{vec}(\bar{X}_1' - \bar{X}_2'))\|_{S_W^{-1}}^2 = \lambda_{r+1} + \lambda_{r+2} + \dots + \lambda_Q. \quad (11)$$

**Projection dans l'espace des lignes ( $\mathbb{R}^J$ ) et des colonnes ( $\mathbb{R}^K$ ).** Les coordonnées des  $K$  lignes (resp. des  $J$  colonnes) de  $X^i$  sur les axes engendrés par les vecteurs  $v_q$ ,  $q = 1, \dots, Q$  (resp.  $u_q$ ,  $q = 1, \dots, Q$ ) sont données respectivement par

$$X^i M v_q = X^i (S_W^R)^{-1} v_q \quad \text{et} \quad (X^i)' D u_q = (X^i)' (S_W^L)^{-1} u_q. \quad (12)$$

### 3 Application: étude des composantes discriminantes

**Données et pré-traitement.** La méthode présentée est appliquée à l'analyse en composantes discriminantes de potentiels évoqués corticaux. Ces signaux EEG ont été acquis dans le cadre du protocole P300 Speller (Farwell & Donchin, 1988) dont le but est d'épeler des lettres en décodant l'activité cérébrale.

Ce dispositif consiste à afficher une matrice  $6 \times 6$  composée de l'alphabet et des chiffres. Le participant souhaitant écrire une lettre doit se concentrer sur la case correspondante dans la matrice, dont les lignes et les colonnes sont successivement et aléatoirement *flashées*. Lorsque la ligne ou la colonne contenant la lettre cible est *flashée*, plusieurs composantes électrophysiologiques sont générées et en particulier une composante visuelle (l'onde négative N1) associée au stimulus visuel, suivie d'une seconde composante liée à la détection du stimulus (l'onde positive P300). Deux classes sont donc considérées: les cibles ( $c = 1$ ) et les non-cibles ( $c = 2$ ). Afin d'épeler une lettre, chaque ligne et chaque colonne est *flashée* 3 fois, ce qui correspond à un total de 36 *flashs* dont 6 sont des cibles. Dans la suite, nous basons nos résultats sur un échantillon de 20 lettres épelées par un participant dans l'expérience menée par Perrin *et al.* (2011).

Chaque essai  $i$  pour  $i = 1, \dots, n$ , avec  $n = 720$ , correspond à une portion du signal enregistré dans l'intervalle de temps  $[-100ms; +900ms]$  où zéro représente l'instant du flash (fréquence d'échantillonnage:  $1000Hz$ ). Une transformation en ondelettes discrètes est appliquée en pré-traitement afin de résumer l'information temporelle. Un filtre de Daubechies D8 est appliqué sur 5 niveaux de décomposition, et seuls les coefficients supérieurs à leur moyenne ont été conservés pour l'analyse. Les données pré-traitées sont donc de la forme  $X_c^i \in \mathbb{R}^{K \times J}$ , où  $K$  est le nombre de coefficients sélectionnés ( $K = 28$ ),  $J$  est le nombre d'électrodes ( $J = 32$ ) et  $c$  précise la classe.

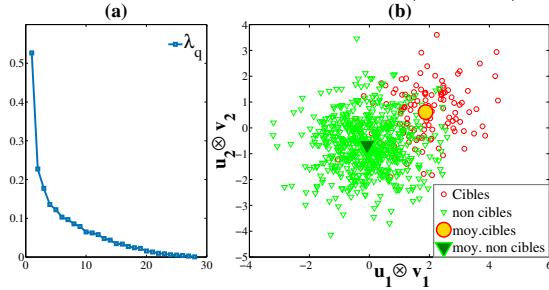


Figure 1: Valeurs propres (a) et représentation des observations et des moyennes dans le premier plan spatio-temporel (b).

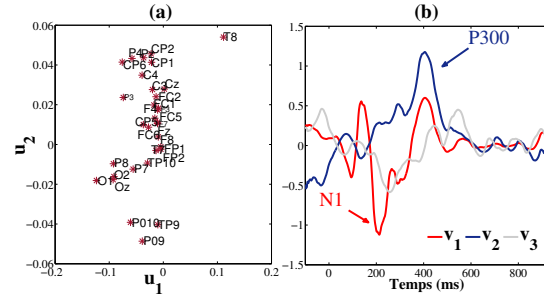


Figure 2: Composantes discriminantes spatiales (a) et temporelles (b) de la différence des moyennes des 2 classes.

**Projection de données dans l'espace spatio-temporel.** La figure 1(a) représente l'éboullis des 28 valeurs propres non nulles. Dans un objectif de réduction de dimension, l'identification d'un "coude" suggère de choisir le sous-espace de projection  $E_r$  de dimension  $r = 4$ . Cette approximation peut être particulièrement utile en classification. La figure 1(b) représente la projection des observations sur le premier plan factoriel engendré par  $u_1 \otimes v_1$  et  $u_2 \otimes v_2$  ainsi que les deux moyennes des cibles et des non-cibles. Cette représentation permet de visualiser la séparation des deux classes dans ce sous-espace et de distinguer les observations atypiques.

**Analyse des composantes discriminantes.** La figure 2(b) représente les trois premières composantes temporelles discriminantes. Elles ont été obtenues par synthèse

d'ondelettes à partir des projections des différences des coefficients moyens entre les classes cible et non-cible sur les axes engendrés par  $v_1$ ,  $v_2$  et  $v_3$ . Les deux principales composantes associées respectivement à  $v_1$  et  $v_2$ , deux vecteurs  $(S_W^R)^{-1}$ -orthogonaux, apparaissent similaires à l'onde N1 et à l'onde P300. Ceci semble donc indiquer une forme de découplage de ces deux composantes. La figure 2(a) représente la projection dans le plan  $(u_1, u_2)$  des différences des moyennes des classes mesurées sur chaque électrode. A partir de cette représentation des groupes distincts d'électrodes peuvent être associés à chacune des composantes précédentes. De plus, on peut détecter des comportements atypiques, comme celui de l'électrode T8. Le comportement singulier de cette électrode située sur la tempe droite est connu (artefacts) et il n'est pas rare de devoir l'exclure des analyses.

## 4 Conclusion

Sous l'hypothèse de séparabilité, la formulation de l'analyse discriminante matricielle binaire comme une SVD de la différence des moyennes des classes offre un cadre générique permettant d'utiliser les propriétés classiques de la SVD. Notamment, la dualité conduit à une analyse conjointe ligne-colonne qui fournit des descripteurs dans les deux domaines simultanément. D'autre part, la réduction de dimension induite peut être utilisée en classification. Dans ce travail nous avons montré la pertinence de cette méthode pour l'analyse spatio-temporelle des signaux EEG. La méthode permet d'extraire et d'analyser distinctement deux composantes (identifiées aux ondes N1 et P300) qui semblent être découplées autant dans le domaine temporel qu'en termes de localisation spatiale.

## Bibliographie

- [1] Guitiérrez, D., et Escalona-Vargas, D.I. (2010), *EEG data classification through signal spatial redistribution and optimized linear discriminants*, Comp.Meth.Prog.Bio., 97, 39-47.
- [2] Akdemir, D. et Gupta, A.K. (2011), *Array variate random variables with multiway Kronecker delta covariance matrix structure*, Journal of Algebraic Statistics, 2(1), 98-113.
- [3] Bijma, F., de Munck, J.C. et Heethaar, R.M. (2005), *The spatiotemporal MEG covariance matrix modeled as a sum of Kronecker products*, NeuroImage, 27, 402-415.
- [4] Friston, K.J., Henson, R.N.A., Phillips, C. et Mattout, J. (2005), *Bayesian estimation of evoked and induced responses*, Human Brain Mapping, 27, 722-735.
- [5] Mahanta, M.S., Aghaei, A.S. et Plataniotis, K.S. (2012), *A Bayes optimal matrix-variate LDA for extraction of spatio-spectral features from EEG signals*, EMBC.
- [6] Spinnato, J., Roubaud M.C., Burle, B. et Torrèsani, B. (2014), *Finding EEG Space-Time-Scale localized features using matrix-based penalized discriminant analysis*, ICASSP.
- [7] Farwell, L. A., Donchin, E. (1988). *Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials*, Electroen. Clin. Neuro., 70(6), 510-523.
- [8] Dutilleul, P. (1999), *The MLE algorithm for the matrix normal distribution*, J. Stat. Comput. Sim., 64(2), 105-123.
- [9] Perrin, M., Maby, E., Bouet, R., Bertrand O. et Mattout, J. (2011), *Detecting and interpreting responses to feedback in BCI*, Graz BCI International Workshop, 116-119.